Wavelet procedures, Lepski method and minimax optimality over Besov spaces

Christophe Chesneau

Wavelet procedures, Lepski method and minimax optimality over Besov spaces

Chesneau Christophe
Laboratoire de Probabilités et Modèles Aléatoires,
CNRS-UMR 7599, Université Paris VI, UFR de Mathématiques,
175 rue de Chevaleret F-75013 Paris, France.
E-mail: christophe.chesneau4@wanadoo.fr
http://www.chesneau-stat.com

13 March 2005

Abstract

The purpose of this paper is to investigate the performance of several adaptive wavelet constructions based on a modification of the original Lepski algorithm. First, we provide a wide class of procedures which have the particularity to attain the asymptotic minimax rate of convergence over certain zones of Besov balls under the $\mathbb{L}^2$ risk. Second, the method developed by Juditsky (1997) is studied. In particular, we show that it is superior to other wavelet estimators in terms of maxiset (and minimax) properties.

Key Words: Maxiset, minimax, adaptive estimation, wavelet block thresholding, Besov spaces.

AMS subject classification: Primary : 62G07 ; Secondary : 62C20.

1 Motivation

Let $\{Y_t^n, \quad t \in [0,1]\}$ be a random process defined by the following stochastic differential equation :

$$dY_t^n = f(t)dt + n^{-\frac{1}{2}}dW_t, \quad t \in [0,1], \quad n \in \mathbb{N}^*, \tag{1.1}$$

where $(W_t)_{t\in[0,1]}$ is a standard Brownian motion and f is an unknown function of $\mathbb{L}^2([0,1])$ that we wish to recover starting from the observations $\left\{\int_0^1 h(t)dY_t^n; \quad h \in \mathbb{L}^2([0,1])\right\}$. This problem has been investigated by many authors under various statistical setting. If we focus our attention on the wavelet methods, the most popular reconstructions are based on the SureShrink and the VisuShrink algorithms. By considering the minimax approach under the mean squared error, we have :

$$\sup_{f\in B^s_{\pi,r}(R)} \mathbb{E}^n_f\left(\int_0^1 |\hat{f}(t) - f(t)|^2 dt\right) \le C\left(\frac{\ln(n)}{n}\right)^{\frac{2s}{1+2s}}, \quad \pi \ge 1.$$

Here, $\hat{f}$ denotes is the estimate of f derived by term-by-term thresholding, $B^s_{\pi,r}(R)$ is the Besov balls (to be defined in Section 2) and $\mathbb{E}^n_f$ is the expectation with respect to the law $\mathbb{P}^n_f$ of the observations. Such procedures are adaptive in the sense where it does not depend on the unknown regularity parameter s. However, they are non-optimal since they come to within a logarithmic factor of the minimax rate of convergence. For more details concerning the procedures and the minimax results described above, see for instance Donoho and Johnstone (1994, 1995).

Over the last decade, numerous efforts have been made to construct adaptive wavelet procedures attaining the exact minimax rate of convergence. Such optimality is achieved for the global thresholding procedure investigated by Kerkyacharian et al. (1996), the version of the Lepski procedure developed by Juditsky (1997), or the block thresholding procedure introduced by Cai (1999).

The present paper can be divided into two parts.

The first part presents several adaptive wavelet procedures based on a modification of the Lepski algorithm which are optimal over $B^s_{\pi,r}(R)$ for $\pi \geq 2$ under the $\mathbb{L}^2$ risk. Initially described by Lepski in a series of papers (see Lepskii (1990, 1991, 1992a,b)), this method has been widely used by many authors in various statistical context. See for instance the following articles : Lepskii and Spokoiny (1995), Juditsky (1997), Lepskii et al. (1997), Lepskii and Levit (1998), Tsybakov (1998), Tsybakov (2000), Birge (2001), Butucea (2000), Kerkyacharian and Picard (2002) and Autin (2005), to name a few. Among other results, we prove that the well-known global thresholding can be viewed as a particular case of the proposed procedures

The second part is devoted to the modification of the Lepski algorithm developed by Juditsky (1997). The idea of this construction is to apply the following term-by-term selection : we keep only the estimators of the wavelet coefficients which are greater than an adaptive threshold parameter determined for each resolution level. If we adopt the minimax point of view, it was shown that they are optimal over $B^s_{\pi,r}(R)$ for $\pi \geq 1$ (including the case $1 \leq \pi < 2$) under the $\mathbb{L}^2$ risk. In this study, we prove that they can outperform the hard thresholding rules in the maxiset sense.

The plan of the paper is organized as follows. Section 2 describes wavelet bases on the interval and the Besov balls. Section 3 sets the main minimax results of some procedures constructed from a modification of Lepski method. The performance of the estimator proposed by Juditsky (1997) is examined in Section 4. Section 5 contains the proofs of Theorems, Propositions and Lemmas.

2 Methodology

2.1 Wavelet bases and Besov balls

We summarize in this subsection the basics on wavelet bases on the unit interval $[0, 1]$.

Let us consider the construction described by Cohen et al. (1993) : We consider ϕ a "father" wavelet of a multiresolution analysis on $\mathbb{R}$ and ψ the associated "mother" wavelet. Assume that $Supp(\phi) = Supp(\psi) = [1-N, N]$ and $\int_{1-N}^{N} \phi(t)dt = 1$, $\int_{1-N}^{N} t^l \psi(t)dt = 0$ for $l = 0, ..., N-1$.

Let us set

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k) \quad \text{and} \quad \psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k).$$

Then there exists an integer τ satisfying $2^\tau \geq 2N$ such that the collection ζ defined by :

$$\zeta = \{\phi_{\tau,k}(.),\ k = 0, ..., 2^\tau - 1;\ \psi_{j,k}(.);\ j \geq \tau,\ k = 0, ..., 2^j - 1\}$$

with an appropriate treatments at the boundaries, is an orthonormal basis of $\mathbb{L}^2([0, 1])$.

Any function f of $\mathbb{L}^2([0, 1])$ can be decomposed on ζ as :

$$f(x) = \sum_{k\in\Delta_\tau} \alpha_{\tau,k}\phi_{\tau,k}(x) + \sum_{j\geq\tau}\sum_{k\in\Delta_j} \beta_{j,k}\psi_{j,k}(x), \qquad x \in [0, 1],$$

where $\alpha_{j,k} = \int_0^1 f(t)\phi_{j,k}(t)dt$, $\beta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt$ and $\Delta_j = \{0, ..., 2^j - 1\}$.

Let us now present the main function spaces of the study :

Definition 2.1 (Besov balls). *Let $N \in \mathbb{N}^*$, $R > 0$, $0 < s < N$, $1 \leq r \leq \infty$ and $1 \leq \pi \leq \infty$. For any function f measurable on $[0, 1]$, we denote the associated N-th order modulus of smoothness as*

$$\rho^N(t, f, \pi) = \sup_{|h|\leq t}\left(\int_{J_{Nh}}\left|\sum_{k=0}^{N}\binom{N}{k}(-1)^k f(u + kh)\right|^\pi du\right)^{\frac{1}{\pi}}$$

where $J_{Nh} = \{x \in [0,1] : \ x + Nh \in [0,1]\}$*. We say that a function* f *of* $\mathbb{L}^{\pi}([0,1])$ *belongs to the Besov balls* $B^s_{\pi,r}(R)$ *if and only if*

$$\left(\int_0^1 \left(\frac{\rho^N(t,f,\pi)}{t^s}\right)^r \frac{1}{t}dt\right)^{\frac{1}{r}} \leq R < \infty$$

with the usual modification if $r = \infty$.

Let us set a well-known result which has been proved by Meyer (1990). Let $0 < s < N$ and $1 \leq \pi \leq \infty$. Then $f \in B^s_{\pi,r}(R)$ if and only if the associated wavelet coefficients $\alpha_{j,k}$ and $\beta_{j,k}$ satisfy :

$$\begin{cases} (2^{\tau(\frac{p}{2}-1)}\sum_{k\in\Delta_\tau}|\alpha_{\tau,k}|^2)^{\frac{1}{2}} + (\sum_{j\geq\tau}(2^{j(s+\frac{1}{2}-\frac{1}{\pi})}(\sum_{k\in\Delta_j}|\beta_{j,k}|^{\pi})^{\frac{1}{\pi}})^r)^{\frac{1}{r}} \leq R < \infty & if \quad r < \infty, \\ (2^{\tau(\frac{p}{2}-1)}\sum_{k\in\Delta_\tau}|\alpha_{\tau,k}|^2)^{\frac{1}{2}} + \sup_{j\geq\tau} 2^{j(s+\frac{1}{2}-\frac{1}{\pi})}(\sum_{k\in\Delta_j}|\beta_{j,k}|^{\pi})^{\frac{1}{\pi}} \leq R < \infty & if \quad r = \infty. \end{cases}$$

3 Some optimal adaptive wavelet procedures

In this section, we propose to study the performance of a wide family of adaptive wavelet procedures constructed from an algorithm close to the Lepski method.

3.1 Wavelet procedures and assumptions

In this subsection, let j_0 be an integer satisfying $2^{j_0} \asymp \ln(n)$, let j_∞ be an integer satisfying $2^{j_\infty} \asymp n$, let m be an integer such that $m \in \{j_0, ..., j_\infty\}$ and let $u = (u_{j,k})_{j,k}$ be positive sequence depending on the data with $u_{j,k} \in \{0,1\}$.

We define the procedures $\hat{f}_{m,u}$ by :

$$\hat{f}_{m,u}(x) = \sum_{j\leq j_0} \hat{\alpha}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{m}\sum_{k\in\Delta_j} \hat{\beta}_{j,k}u_{j,k}\psi_{j,k}(x), \qquad x \in [0,1]. \tag{3.1}$$

where $\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(t)dY_t^n$ and $\hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(t)dY_t^n$.

Such constructions includes several well-known procedures as the linear procedures, the hard thresholding procedures, the global thresholding procedures and the block thresholding procedures. They can be viewed as a generalization of the μ-thresholding procedures developed by Autin (2005).

Let us now consider the following assumption :

Assumption H**.** There exists a constant $C > 0$ such that the random positive sequence $u = (u_{j,k})_{j,k}$ satisfies :

$$(\sum_{k\in\Delta_j} |\hat{\beta}_{j,k}|^2(1-u_{j,k}))^{\frac{1}{2}} \leq Cn^{-\frac{1}{2}}2^{\frac{j}{2}}, \quad j \in \{j_0, ..., j_\infty\}. \tag{3.2}$$

3.2 Upper bounds

The notation $a \asymp b$ means : there exist two constants $C > 0$ and $c > 0$ such that $cb \leq a \leq Cb$. The notations $a \wedge b$ and $a \vee b$ mean respectively : $\min(a,b)$ and $\max(a,b)$.

Theorem 3.1 below is standard.

Theorem 3.1. *Let $R > 0$. Let $\hat{f}_{j_s,u}$ be the procedure described in (4.1) where j_s is an integer satisfying :*

$$2^{j_s} \asymp n^{\frac{1}{1+2s}}. \tag{3.3}$$

Then, under the assumption H, for $0 < s < N$, $\pi \geq 2$, and $r \geq 1$ there exists a constant $C > 0$ such that :

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}_{j_s,u} - f\|_2^2) \leq C n^{-\frac{2s}{1+2s}}, \qquad n \geq n_0,$$

for n_0 large enough.

Due to the presence of the regularity parameter s, the procedure $\hat{f}_{j_s,u}$ is not adaptive.

In order to determine adaptive procedures which attain the previous rate of convergence, the main idea of Theorem 3.2 below is the following : we estimate the integer j_s by a random integer $\hat{\jmath}$ depending on the data and we consider the procedure $\hat{f}_{\hat{\jmath},u}$. Such point of view is inspired by the original Lepski algorithm developed in Lepskii (1990).

Theorem 3.2. *Let $R > 0$. Let $\hat{f}_{\hat{\jmath},u}$ be the procedure described in (4.1) where $\hat{\jmath}$ is the following random integer :*

$$\hat{\jmath} = \min\{l \in \{j_0, ..., j_\infty\}; \quad \|\hat{f}_{l,u} - \hat{f}_{l-1,u}\|_2 \leq \kappa_1 2^{\frac{l}{2}} n^{-\frac{1}{2}}\} - 1, \tag{3.4}$$

and κ_1 is the real number satisfying $\kappa_1^2 - 8\ln(2^{-1}\kappa_1) \geq 20$. (Let us precise that we have adopt the convention $\hat{f}_{j_0-1,u} = 0$).

Then under the assumption H, for $0 < s < N$, $\pi \geq 2$ and $r \geq 1$ there exists a constant $C > 0$ such that :

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}_{\hat{\jmath},u} - f\|_2^2) \leq C n^{-\frac{2s}{1+2s}}, \qquad n \geq n_0,$$

for n_0 large enough.

See below two equivalent expressions of the random integer $\hat{\jmath}$:

$$\hat{\jmath} = \begin{cases} \min\{l \in \{j_0, ..., j_\infty\}; & (\sum_{k \in \Delta_l} |\hat{\beta}_{l,k}|^2 u_{l,k})^{\frac{1}{2}} \leq \kappa_1 2^{\frac{l}{2}} n^{-\frac{1}{2}}\} - 1, \\ \max\{l \in \{j_0, ..., j_\infty\}; & (\sum_{k \in \Delta_l} |\hat{\beta}_{l,k}|^2 u_{l,k})^{\frac{1}{2}} > \kappa_1 2^{\frac{l}{2}} n^{-\frac{1}{2}}\}. \end{cases}$$

Let us mention that the upper bound above is without logarithmic factor contrary to the rate attained by the hard thresholding procedures under the same circumstances. See for instance Donoho and Johnstone (1995). Let us recall than the hard thresholding procedure can be defined by $\hat{f}_{j_\infty,u}$ (see (4.1)) where :

$$u_{j,k} = 1_{\left\{|\hat{\beta}_{j,k}| \geq \kappa_2 \sqrt{\frac{\ln(n)}{n}}\right\}}, \tag{3.5}$$

where $\kappa_2 \geq 4$.

For sake of simplicity, such procedures will be denoted $\hat{f}^h$.

3.3 On the assumption H

Now, we propose to examine the connections which exist between the constructions $\hat{f}_{\hat{\jmath},u}$ and some well-known wavelet procedures by considering several sequences u satisfying the assumption H.

1. Let $u = (u_{j,k})_{j,k}$ be the positive sequence defined by :
$$u_{j,k} = 1.$$
Then the procedure $\hat{f}_{\hat{\jmath},u}$ is close to the 'true' Lepski procedure introduced by Lepskii (1990).
2. Let $u = (u_{j,k})_{j,k}$ be the positive sequence defined by :
$$u_{j,k} = 1_{\left\{(2^{-j}\sum_{l\in\Delta_j}|\hat{\beta}_{j,l}|^2)^{\frac{1}{2}} \geq cn^{-\frac{1}{2}}\right\}},$$
where c denotes a positive constant. If $c \geq \kappa_1$ then the procedure $\hat{f}_{\hat{\jmath},u}$ becomes the global thresholding procedure (defined by $\hat{f}_{j_\infty,u}$). For further details concerning such construction, see Kerkyacharian et al. (1996).
3. Let $u = (u_{j,k})_{j,k}$ be the positive sequence defined by :
$$u_{j,k} = 1_{\left\{|\hat{\beta}_{j,k}| \geq cn^{-\frac{1}{2}}\right\}}$$
where c denotes a positive constant. For such a choice, let us notice that $\hat{f}_{\hat{\jmath},u}$ becomes a hybrid version of the hard thresholding procedure. Let us just remark that no logarithmic factor appears in the threshold.
4. Finally, let us precise that the BlockShrink sequence described by Cai (1999) satisfies the assumption H.

Remark 3.1. *It is important to mention that the adaptive procedures described above have several drawbacks. Among them, due to the definition of $\hat{\jmath}$, it seems difficult to exhibit a 'good' rate of convergence for $\hat{f}_{\hat{\jmath},u}$ over $B^s_{\pi,\infty}(R)$ in the case where $1 \leq \pi < 2$.*

For instance, if we consider $u = 1$ and the statistical problem as given in (1.1) then we have for $1 \leq \pi < 2$:
$$\sup_{f\in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}_{j_s,1} - f\|_2^2) \geq cn^{-\frac{2s'}{1+2s'}} > cn^{-\frac{2s}{1+2s}} \geq c \inf_{\tilde{f}} \sup_{f\in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\tilde{f} - f\|_2^2)$$
where $s' = s + \frac{1}{2} - \frac{1}{\pi}$. For a general proof, we refer the reader to Nemirovskii (1986). Thus, if we estimate j_s by a random integer then it is natural that the obtained procedure does not attain the minimax rate of convergence for $1 \leq \pi < 2$.

An alternative is developed in the section below.

4 Hard thresholding procedures with random threshold

Delyon and Juditsky (1996) have shown that the non-adaptive procedure defined by $\hat{f}_{j_1,u}$ (see (4.1)) where :
$$u_{j,k} = 1_{\left\{|\hat{\beta}_{j,k}| \geq \kappa_3\sqrt{\frac{(j-j_s)_+}{n}}\right\}},$$
$\kappa_3 \geq 4\sqrt{\ln(2)}$, j_1 is an integer such that $2^{j_1} \asymp \frac{n}{\ln(n)}$ and j_s is an integer such that $2^{j_s} \asymp n^{\frac{1}{1+2s}}$ can be optimal over $B^s_{\pi,\infty}(R)$ in the case where $\pi \geq 1$ for numerous statistical models.

Starting from this result and using Lepski algorithm, an adaptive construction was developed by Juditsky (1997). The main idea is to estimate an adaptive threshold parameter for each resolution level.

We propose to discuss and compare the performance of this procedure with other well-known constructions via the minimax point of view and the maxiset approach.

4.1 The Juditsky procedure

For all $l \in \{0, ..., j\}$, we set :

$$\tilde{\beta}_{j,k}(l) = \hat{\beta}_{j,k} 1_{\left\{|\hat{\beta}_{j,k}| \geq \kappa_3 \sqrt{\frac{l}{n}}\right\}} \quad \text{and} \quad M_j(l) = 2^{j-l} n^{-1}$$

where $\kappa_3 \geq 4\sqrt{\ln(2)}$.

We define the random integer $\hat{l}_j$ by :

$$\hat{l}_j = \begin{cases} \min\left\{l \in \{1, ..., j\}; \quad \exists r \in \{0, ..., l-1\}, \quad \sum_{k \in \Delta_j} |\tilde{\beta}_{j,k}(l) - \tilde{\beta}_{j,k}(r)|^2 \geq 9M_j(r)\right\} - 1, \\ \\ j \quad \text{if} \quad \sum_{k \in \Delta_j} |\tilde{\beta}_{j,k}(l) - \tilde{\beta}_{j,k}(r)|^2 \leq 9M_j(r), \quad \forall r < l \leq j. \end{cases}$$

A slight modification of the construction described by Juditsky (1997) can be defined by :

$$\hat{f}^J(x) = \sum_{j \leq j_0} \hat{\alpha}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j_0 \leq j \leq j_1} \sum_{k \in \Delta_j} \tilde{\beta}_{j,k}(\hat{l}_j) \psi_{j,k}(x), \quad x \in [0,1], \tag{4.1}$$

where j_0 is an integer such that $2^{j_0} \asymp \ln(n)^3$ and j_1 is an integer such that $2^{j_1} \asymp \frac{n}{\ln(n)}$.

Remark 4.1. *Let us precise that the original procedure is defined for j_0 such that $2^{j_0} \asymp n^{\frac{1}{1+2N}}$. After calculus, one can chose j_0 such that $2^{j_0} \asymp \ln(n)^3$ without deteriorating the performance of the procedure.*

Theorem below is a $\mathbb{L}^2$-version of Juditsky (1997, Theorem 1). It shows that the procedure described above enjoys good minimax properties over Besov balls under the $\mathbb{L}^2$ risk.

Theorem 4.1 (Juditsky (1997)). *Let $R > 0$. Let $\hat{f}^J$ be the procedure described in (4.1). Then for $(\pi^{-1} - 2^{-1})_+ < s < N$, $\pi \geq 1$ and $r \geq 1$ there exists a constant $C > 0$ such that :*

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}^J - f\|_2^2) \leq C n^{-\frac{2s}{1+2s}}, \qquad n \geq n_0,$$

for n_0 large enough.

It is important to mention that for the case $2 > \pi \geq 1$, the procedure $\hat{f}^J$ is without logarithmic factor (and can be optimal for several statistical models) contrary to the rate of convergence attained by :

1. The hard thresholding procedure (see for instance Donoho et al. (1996)),
3. The kernel estimator developed by Lepskii et al. (1997),
2. The BlockShrink procedure introduced by Cai (1999).

4.2 Maxiset properties of $\hat{f}^J$

Now, let us investigate the maxiset properties of $\hat{f}^J$. For any estimator $\tilde{f}$ and any positive sequence c_n, we define the maxiset of $\tilde{f}$ at the rate of convergence c_n by :

$$\mathcal{M}ax(\tilde{f}, c_n)(D) = \left\{f \in \mathbb{L}^2([0,1]); \quad \sup_{n>0} c_n^{-1} \mathbb{E}_f(\|\tilde{f} - f\|_2^2) \leq D < \infty\right\}$$

Proposition 4.1 below compares the maxiset associated to $\hat{f}^J$ and $\hat{f}^h$ at the rate $n^{-\alpha}$, $\alpha \in]0,1[$.

Proposition 4.1. *Let $\hat{f}^h$ be the hard thresholding procedure (see 3.5) and let $\hat{f}^J$ be the procedure defined by (4.1). Then for any $\alpha \in]0,1[$ we have the following inclusion :*

$$\mathcal{M}ax(\hat{f}^h, n^{-\alpha}) \subseteq \mathcal{M}ax(\hat{f}^J, n^{-\alpha}), \qquad n \geq n_0,$$

for n_0 large enough. In other words, $\hat{f}^J$ is better in the maxiset sense than $\hat{f}^h$.

Let us notice that we have a similar maxiset result for the hard thresholding and the BlockShrink constructions (see for instance Autin (2005) or Chesneau (2006)).

Remark 4.2. *It is easy to show that the previous inclusion holds if we consider the rate of convergence $(\frac{\ln(n)}{n})^\alpha$, $\alpha \in]0,1[$.*

Conclusion : The Lepski method combined with wavelet bases can provide procedures which enjoy optimal properties in the minimax and maxiset senses.

5 Appendix : proofs of Theorems

Proof of Theorem 3.1. Suppose that the parameters s, π and r of the Besov spaces $B^s_{\pi,r}(R)$ satisfy $s > 0$, $\pi \geq 2$ and $r \geq 1$. Firstly, let us investigate the rate of convergence attains by the procedure $\hat{f}_{j_s,1}$.

The $\mathbb{L}^2$ risk of $\hat{f}_{j_s,1}$ can be bounded by a sum of three components :

$$\mathbb{E}^n_f(\|\hat{f}_{j_s,1} - f\|_2^2) \leq 3(F_1 + F_2 + F_3) \tag{5.1}$$

where

$$F_1 = \sum_{k \in \Delta_{j_0}} \mathbb{E}^n_f(|\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2), \quad F_2 = \sum_{j=j_0}^{j_s} \sum_{k \in \Delta_j} \mathbb{E}^n_f(|\hat{\beta}_{j,k} - \beta_{j,k}|^2) \quad F_3 = \sum_{j=j_s+1}^{\infty} \sum_{k \in \Delta_j} |\beta_{j,k}|^2.$$

Since $\hat{\alpha}_{j,k} - \alpha_{j,k} \sim \hat{\beta}_{j,k} - \beta_{j,k} \sim \mathcal{N}(0, n^{-1})$ and $f \in B^s_{\pi,r}(R) \subseteq B^s_{2,\infty}(R)$ (since $\pi \geq 2$ and $r \geq 1$), it is easy to see that :

$$F_1 + F_2 + F_3 \leq C(2^{j_0} n^{-1} + n^{-1} \sum_{j=j_0}^{j_s} 2^j + 2^{-2 j_s s}) \leq C n^{-\frac{2s}{1+2s}}$$

for n large enough. So :

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}_{j_s,1} - f\|_2^2) \leq C n^{-\frac{2s}{1+2s}} \tag{5.2}$$

for n large enough. The assumption H yields :

$$\|\hat{f}_{j_s,u} - \hat{f}_{j_s,1}\|_2^2 = \sum_{j=j_0}^{j_s} \sum_{k \in \Delta_j} |\hat{\beta}_{j,k}|^2 (1 - u_{j,k}) \leq C 2^{j_s} n^{-1} \leq C n^{-\frac{2s}{1+2s}}. \tag{5.3}$$

Combining (5.1)-(5.3), one gets :

$$\mathbb{E}^n_f(\|\hat{f}_{j_s,u} - f\|_2^2) \leq 2(\mathbb{E}^n_f(\|\hat{f}_{j_s,1} - f\|_2^2) + \mathbb{E}^n_f(\|\hat{f}_{j_s,u} - f_{j_s,1}\|_2^2)) \leq C n^{-\frac{2s}{1+2s}}$$

for n large enough. This ends the proof of Theorem 3.1. $\square$

Proof of Theorem 3.2. Suppose that the parameters s, π and r of the Besov space $B^s_{\pi,r}(R)$ satisfy $s > 0$, $\pi \geq 2$ and $r \geq 1$. Firstly, let us show that there exists a constant $C > 0$ satisfying :

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{E}^n_f(\|\hat{f}_{\hat{j},u} - \hat{f}_{j_s,u}\|_2^2) \leq C n^{-\frac{2s}{1+2s}} \tag{5.4}$$

for n large enough.

We have :

$$V = \mathbb{E}^n_f(\|\hat{f}_{\hat{j},u} - \hat{f}_{j_s,u}\|_2^2) \leq 2(V_1 + V_2) \tag{5.5}$$

where we have set :

$$V_1 = \mathbb{E}^n_f(\|(\hat{f}_{\hat{j},u} - \hat{f}_{j_s,u})1_{\{\hat{j}>j_s\}}\|_2^2) \quad \text{and} \quad V_2 = \mathbb{E}^n_f(\|(\hat{f}_{\hat{j},u} - \hat{f}_{j_s,u})1_{\{\hat{j}<j_s\}}\|_2^2).$$

Let us analyze each term V_i, i=1,2 in turn.

The upper bounds for V_1. Using the equality $max(u_{j,k}, 1 - u_{j,k}) = 1$, the fact that the integers j_s and $\hat{j}$ belong to $\{j_0, ..., j_\infty\}$, and the elementary decomposition $\hat{\beta}_{j,k} = (\hat{\beta}_{j,k} - \beta_{j,k}) + \beta_{j,k}$, we observe that :

$$V_1 \leq C \sum_{j=j_0}^{j_\infty} \sum_{k \in \Delta_j} \mathbb{E}^n_f(|\hat{\beta}_{j,k}|^2 1_{\{\hat{j}>j_s\}}) \leq C(H_1 + H_2)$$

where

$$H_1 = \sum_{j=j_0}^{j_\infty} \sum_{k \in \Delta_j} \mathbb{E}^n_f(|\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{\hat{j}>j_s\}}) \text{ and } H_2 = \sum_{j=j_0}^{j_\infty} \sum_{k \in \Delta_j} |\beta_{j,k}|^2 \mathbb{P}^n_f(\hat{j} > j_s). \tag{5.6}$$

In order to bound the terms H_1 and H_2, let us consider the following lemma which will be proved at the end of the present proof.

Lemma 5.1. *If $s > 0$, $\pi \geq 2$ and $r \geq 1$ then there exists a constant $C > 0$ such that :*

$$\sup_{f \in B^s_{\pi,r}(R)} \mathbb{P}^n_f(\hat{j} > j_s) \leq C n^{-2}.$$

It follows from the Cauchy-Schwartz inequality and Lemma 5.1 that :

$$\mathbb{E}^n_f(|\hat{\beta}_{j,k} - \beta_{j,k}|^2 1_{\{\hat{j}>j_s\}}) \leq \mathbb{E}^n_f(|\hat{\beta}_{j,k} - \beta_{j,k}|^4)^{\frac{1}{2}} (\sup_{f \in B^s_{\pi,r}(R)} \mathbb{P}^n_f(\hat{j} > j_s))^{\frac{1}{2}} \leq C n^{-2}.$$

By virtue of the previous inequality and the definition of j_∞, it comes :

$$H_1 \leq C n^{-2} \sum_{j=j_0}^{j_\infty} 2^j \leq C 2^{j_\infty} n^{-2} \leq C n^{-1} \leq C n^{-\frac{2s}{1+2s}}$$

for n large enough.

Lemma 5.1 allows us to dominate H_2 :

$$H_2 \leq C \|f\|_2^2 \sup_{f \in B^s_{\pi,r}(R)} \mathbb{P}^n_f(\hat{j} > j_s) \leq C n^{-\frac{2s}{1+2s}}$$

for κ large enough. We deduce from the upper bounds of H_1 and H_2 that there exists a constant $C > 0$ such that :

$$V_1 \leq Cn^{-\frac{2s}{1+2s}} \tag{5.7}$$

for κ large enough.

The upper bound for V_2. By definition of $\hat{j}$, for all $j \geq \hat{j}+1$ we have :

$$\sum_{k\in\Delta_j} |\hat{\beta}_{j,k}|^2 u_{j,k} \leq \kappa_1^2 2^j n^{-1},$$

so :

$$V_2 \leq C\mathbb{E}_f^n(\sum_{j=\hat{j}+1}^{j_s}\sum_{k\in\Delta_j} |\hat{\beta}_{j,k}|^2 u_{j,k} 1_{\{\hat{j}<j_s\}}) \leq Cn^{-1}\sum_{j=j_0}^{j_s} 2^j \leq Cn^{-1}2^{j_s} \leq Cn^{-\frac{2s}{1+2s}}. \tag{5.8}$$

Putting (5.5) and (5.7)-(5.8) together, we deduce that the inequality (5.4) holds. Combining this result with Theorem 3.1, we see that :

$$\mathbb{E}_f^n(\|\hat{f}_{\hat{j},u} - f\|_2^2) \leq C(\mathbb{E}_f^n(\|\hat{f}_{\hat{j},u} - \hat{f}_{j_s,u}\|_2^2) + \mathbb{E}_f^n(\|\hat{f}_{j_s,u} - f\|_2^2)) \leq Cn^{-\frac{2s}{1+2s}}$$

for n large enough. This completes the proof of Theorem 3.2. $\square$

Proof of Lemma 5.1. By definition of $\hat{j}$, if $\hat{j} > j_s$, then there exists $j_* \in \{j_s+1, ..., j_\infty\}$ such that :

$$\sum_{k\in\Delta_{j_*}} |\hat{\beta}_{j_*,k}|^2 \geq \sum_{k\in\Delta_{j_*}} |\hat{\beta}_{j_*,k}|^2 u_{j_*,k} > \kappa_1^2 2^{j_*} n^{-1}.$$

Since $f \in B^s_{\pi,r}(R) \subseteq B^s_{2,\infty}(R)$ ($\pi \geq 2$ and $1 \leq r \leq \infty$), for any $j > j_s$ (where j_s an integer such that $2^{j_s} \geq cn^{\frac{1}{1+2s}}$ with c is a constant suitably chosen) we have :

$$\sum_{k\in\Delta_j} |\beta_{j,k}|^2 \leq R2^{-j_s(2s+1)}2^j \leq R2^{-j_s(2s+1)}2^j \leq \frac{\kappa_1^2}{4}2^j n^{-1}.$$

It follows from the l_2 Minkowski inequality that :

$$(\sum_{k\in\Delta_{j_*}} |\hat{\beta}_{j_*,k} - \beta_{j_*,k}|^2)^{\frac{1}{2}} \geq (\sum_{k\in\Delta_{j_*}} |\hat{\beta}_{j_*,k}|^2)^{\frac{1}{2}} - (\sum_{k\in\Delta_{j_*}} |\beta_{j_*,k}|^2)^{\frac{1}{2}} > \frac{\kappa_1}{2}2^{\frac{j_*}{2}}n^{-\frac{1}{2}}.$$

Using a tail probability of the $\chi^2_{2^j}$ which appears in Cai and Silverman (2001, Subsection 5.3), the choice of κ_1 and the fact that $2^{j_0} \asymp \ln(n)$, one gets :

$$\begin{aligned} \mathbb{P}_f^n((2^{-j_*}\sum_{k\in\Delta_{j_*}} |\hat{\beta}_{j_*,k} - \beta_{j_*,k}|^2)^{\frac{1}{2}} \geq \frac{\kappa_1}{2}n^{-\frac{1}{2}}) &\leq C2^{-\frac{j_*}{2}}\exp(-2^{j_*-1}(4^{-1}\kappa_1^2 - 1 - 2\ln(2^{-1}\kappa_1))) \\ &\leq Cn^{-2}2^{-\frac{j_*}{2}}. \end{aligned}$$

It follows that :

$$\mathbb{P}_f^n(\hat{j} > j_s) \leq \sum_{j=j_s+1}^{j_\infty} \mathbb{P}_f^n((2^{-j}\sum_{k\in\Delta_j} |\hat{\beta}_{j,k} - \beta_{j,k}|^2)^{\frac{1}{2}} \geq \frac{\kappa_1}{2}n^{-\frac{1}{2}}) \leq Cn^{-2}\sum_{j=j_s+1}^{j_\infty} 2^{-\frac{j}{2}} \leq Cn^{-2}.$$

The proof of Lemma 5.1 is thus complete. $\square$

Proof of Proposition 4.1. The proof is a direct consequence of the construction of the procedure $\hat{f}^J$. By a slight modification of the Juditsky (1997, Proposition 1), we can show that there exists a constant $C > 0$ such that :

$$\sum_{k\in\Delta_j} \mathbb{E}^n_f(|\tilde{\beta}_{j,k}(\hat{l}) - \beta_{j,k}|^2) \leq C(\min_{\lambda\geq 0} \mathcal{R}_j(\lambda) + \frac{\ln(n)^5}{n}), \qquad j \in \{j_0, ..., j_1\},$$

where

$$\mathcal{R}_j(\lambda) = 2^{j-\lambda}n^{-1} + \sum_{k\in\Delta_j} (\min(\kappa_3\sqrt{\frac{\lambda}{n}}, |\beta_{j,k}|))^2.$$

Now, let us remark that for all $\lambda \geq 0$, the Markov inequality gives us :

$$\begin{aligned} \kappa_3^2\lambda n^{-1}\sum_{k\in\Delta_j} 1_{\left\{|\beta_{j,k}|>\kappa_3\sqrt{\frac{\lambda}{n}}\right\}} &= \kappa_3^2\lambda n^{-1}\sum_{m\in\mathbb{N}}\sum_{k\in\Delta_j} 1_{\left\{\kappa_3 2^{m+1}\sqrt{\frac{\lambda}{n}}\geq|\beta_{j,k}|>\kappa_3 2^m\sqrt{\frac{\lambda}{n}}\right\}} \\ &\leq \sum_{m\in\mathbb{N}} 2^{-2m}\sum_{k\in\Delta_j} |\beta_{j,k}|^2 1_{\left\{|\beta_{j,k}|\leq\kappa_3 2^{m+1}\sqrt{\frac{\lambda}{n}}\right\}}. \end{aligned}$$

So, we deduce that

$$\begin{aligned} \mathcal{R}_j(\lambda) &= 2^{j-\lambda}n^{-1} + \kappa_3^2\lambda n^{-1}\sum_{k\in\Delta_j} 1_{\left\{|\beta_{j,k}|>\kappa_3\sqrt{\frac{\lambda}{n}}\right\}} + \sum_{k\in\Delta_j} |\beta_{j,k}|^2 1_{\left\{|\beta_{j,k}|\leq\kappa_3\sqrt{\frac{\lambda}{n}}\right\}}. \\ &\leq 2^{j-\lambda}n^{-1} + \sum_{m\in\mathbb{N}} 2^{-2m}\sum_{k\in\Delta_j} |\beta_{j,k}|^2 1_{\left\{|\beta_{j,k}|\leq\kappa_3 2^{m+1}\sqrt{\frac{\lambda}{n}}\right\}}. \end{aligned}$$

Let us recall that if $f \in \mathcal{M}ax(\hat{f}^h, n^{-\alpha})$ for n_0 large enough then :

- $$\sup_{u>0} \eta(u)^{-\alpha}\sum_j\sum_{k\in\Delta_j} |\beta_{j,k}|^2 1_{\{|\beta_{j,k}|\leq u\}} \leq R < \infty,$$

where η is the continuous non decreasing function defined by :

$$\eta(u) = \begin{cases} u\ln((u\wedge\upsilon)^{-1})^{-\frac{1}{2}}, & u > 0, \\ 0, & u = 0, \end{cases}$$

and υ is a real number such that $0 < \upsilon \leq \exp(-\frac{1}{2(1-\alpha)} + 1)$.

- $$\sup_{n>0} n^\alpha \sum_{j\geq j_1+1}\sum_{k\in\Delta_j} |\beta_{j,k}|^2 \leq C < \infty.$$

For more details concerning the previous result, see for instance Autin (2005) or Chesneau (2006).

Combining all the previous results and using Kerkyacharian et al. (2005, Lemma 2), if $f \in \mathcal{M}ax(\hat{f}^h, n^{-\alpha})$ then we have :

$$\begin{aligned} \sum_{j=j_0}^{j_1} \mathcal{R}_j(\kappa_o^2\ln(n)) &\leq C(n^{-1}2^{j_\infty}2^{-\kappa_o^2\ln(n)} + \sum_{m\in\mathbb{N}} 2^{-2m}\eta(\sqrt{\frac{\ln(n)}{n}}C\kappa_o\kappa_3 2^{m+1})^{2\alpha}) \\ &\leq C(2^{-\kappa_o^2\ln(n)} + (\frac{\ln(n)}{n})^\alpha(\ln((C\kappa_o\kappa_3\sqrt{\frac{\ln(n)}{n}}\wedge\upsilon)^{-1}))^{-\alpha}) \\ &\leq Cn^{-\alpha}. \end{aligned}$$

for κ_0 large enough. Thus, if $f \in \mathcal{M}ax(\hat{f}^h, n^{-\alpha})$ then it follows from a decomposition similar to Theorem 3.1 and the previous inequality that :

$$\begin{aligned}
\mathbb{E}^n_f(\|\hat{f}^J - f\|_2^2) &\leq C(\sum_{k\in\Delta_{j_0}} \mathbb{E}^n_f(|\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2) + \sum_{j=j_0}^{j_1} \mathcal{R}_j(\kappa_o^2 \ln(n)) + j_1 \frac{\ln(n)^5}{n} + \sum_{j\geq j_1+1}\sum_{k\in\Delta_j} |\beta_{j,k}|^2) \\
&\leq C(\ln(n)^3 n^{-1} + n^{-\alpha} + j_1 \frac{\ln(n)^5}{n} + n^{-\alpha}) \leq C n^{-\alpha}.
\end{aligned}$$

for n and κ_o large enough. We conclude that $\mathcal{M}ax(\hat{f}^h, n^{-\alpha}) \subseteq \mathcal{M}ax(\hat{f}^J, n^{-\alpha})$. □

Références

AUTIN, F. (2005). *Point de vue maxiset en estimation non paramétrique*. PhD Thesis, Laboratoire de probabilités et modèles aléatoires, Université Paris VII.

BIRGE, L. (2001). An alternative point of view on lepski's method. *State of the art of probability and statistics (Leiden, 1999) (ed. Monogr. IMS Lectures notes)*, (36) :113–133.

BUTUCEA, C. (2000). Two adaptive rates of convergence in pointwise density estimation. *Mathematical Methods of Statistics*, 9(1) :39–64.

CAI, T. (1999). Adaptative wavelet estimation : a block thresholding and orakle inequality approach. *Ann. Stat.*, 27 :898–924.

CAI, T. and SILVERMAN, B. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya*, (63) :127–148.

CHESNEAU, C. (2006). On block thresholding in nonparametric estimation. *Preprint Laboratoire de Probabilités et Modèles Aléatoires, Paris VI. This article can be downloaded on* *http://www.chesneau-stat.com*.

COHEN, A., DAUBECHIES, I., JAWERTH, B., and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1) :54–81.

DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3 :215–228.

DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81 :425–455.

DONOHO, D. and JOHNSTONE, I. (1995). Adaptating to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432) :1200–1224.

DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G., and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539.

JUDITSKY, A. (1997). Wavelet estimators : adapting to unknown smoothness. *Math. Methods of Statistics*, (1) :1–20.

KERKYACHARIAN, G. and PICARD, D. (2002). Minimax or maxisets ? *Bernouilli*, 8(2) :219–253.

KERKYACHARIAN, G., PICARD, D., and RAIMONDO, M. (2005). Adaptive boxcar deconvolution on full lebesgue measure sets. Technical report, Department of Statistics.

Kerkyacharian, G., Picard, D., and Tribouley, K. (1996). Lp adaptive density estimation. *Bernouilli*, 2 :229–247.

Lepskii, O. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.*, 35 :454–466.

Lepskii, O. (1991). Asymptotically minimax adaptive estimation i : Upper bounds. optimally adaptive estimates. *Theory Probab. Appl.*, 36 :682–697.

Lepskii, O. (1992a). Asymptotically minimax adaptive estimation ii : Statistical model without optimal adaptation. adaptive estimators. *Theory Probab. Appl.*, 37 :433–468.

Lepskii, O. (1992b). On problems of adaptiveestimation in white gaussian noise. *Theory Probab. Appl.*, 12 :87–106.

Lepskii, O. and Levit, B. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, 7 :123–156.

Lepskii, O., Mammen, E., and Spokoiny, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness : An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25 :929–947.

Lepskii, O. and Spokoiny, V. (1995). Local adaptation to inhomogeneous smoothness : resolution level. *Math. Methods Statist.*, 4 :239–258.

Meyer, Y. (1990). *Ondelettes et Opérateurs*. Hermann, Paris.

Nemirovskii, A. S. (1986). Nonparametric estimation of smooth regression function. *Journal of Computer and Systems Sciences*, 23(6) :1–11.

Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann.Statist.*, 28(1) :298–335.

Tsybakov, A. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on sobolev classes. *Ann. Statist.*, 26 :2420–2469.

Tsybakov, A. (2000). Adaptive estimation for inverse problems : a logarithmic effect in l2. *C.R. Acad. Sci. Paris, ser. 1, t.330*, pp. 835–840.